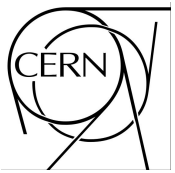


The obsolescence of Information and Information Systems

The Digital Memory project

Jean-Yves Le Meur - Information Technology - 2018

World Wide Web

The WorldWideWeb (W3) is a wide-area hypermedia information retrieva
universe of documents.

Everything there is online about W3 is linked directly or indirectly to this
project, Mailing lists , Policy , November's W3 news , Frequently Asked Q

What's out there?
Pointers to the world's online information, subjects , W3 servers, etc
Help
on the browser you are using
Software Products
A list of W3 project components and their current state. (e.g. Line M
robot , Library)
Technical
Details of protocols, formats, program internals etc
Bibliography
Paper documentation on W3 and references.
People
A list of some people involved in the project.
History
A summary of the history of the project.
How can I help ?
If you would like to support the web..
Getting code
Getting the code by anonymous FTP , etc.

Digital preservation at CERN: step by step

A short history of commitments

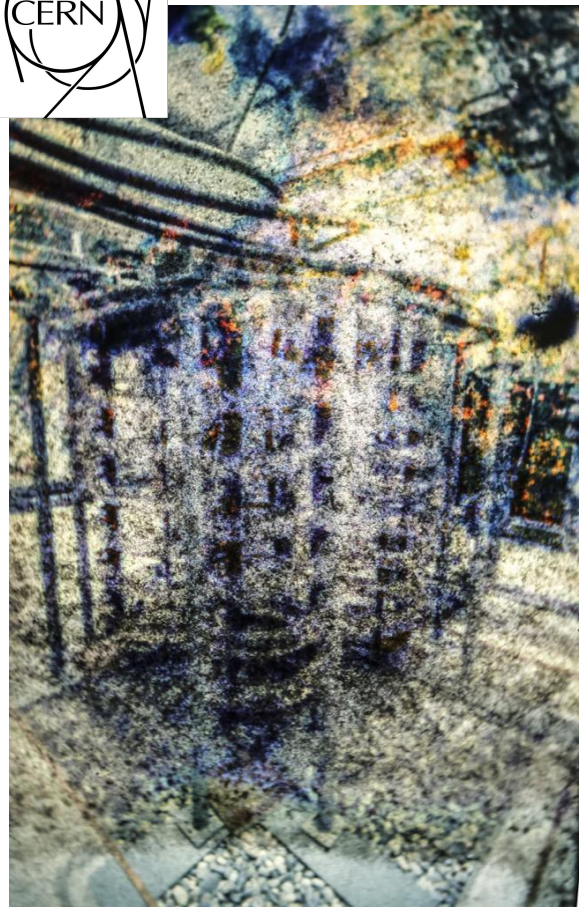
- 1997 — Long Term Electronic Archiving working group
- 2004 — LTEA advices approved by Directorate
- 2006 — Integrated Project Support Study fosters an e-archival policy
- 2009 — HEP Data preservation launched
- 2016 — **Digital Memory** project started

The urgency of safeguarding multimedia

audio


video/film


photo



Starting in troubled waters

Inventory ?
- ➢ no numbers

Carriers ?
- ➢ no uniformity

Catalog ?
- ➢ no metadata

Funding ?

Knowledge
- ➢ no experts

Formats ?
- ➢ no standards

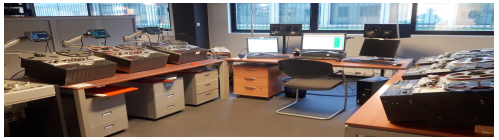
Memoriav partnership for **Videotapes** digitization

CERN Council starts to fund **Audio** conversion

CERN IT supports full **AV** & **Photos** scanning

2016 — 2017

The Digitization



Audio recordings

Speeches of **official committees** since 1956: closed access to Archive & Translation services

→ 7’500 tapes & cassettes converted to **wav** uncompressed (48kHz&16 bit) + ogg and mp3 (320 kbps) files

Speech to text ?

Video recordings

Footages, documentaries and seminars; 13 different carrier types; everywhere on CERN site: unclear access rights

→ 5’700 carriers converted to **mkv/ffv1** (4K,2K or SD) & mov and mp4 files : with ~10% of “dead/damaged” supports

Slides & Negatives

120’000 B&W photos scanned in 2014, **300’000 color pictures** captured by photoLab:organized in 13’000 albums

→ converted to **TIFF** (48 bit) and JPEG (24 bit) at 4800 ppi for 24x36mm

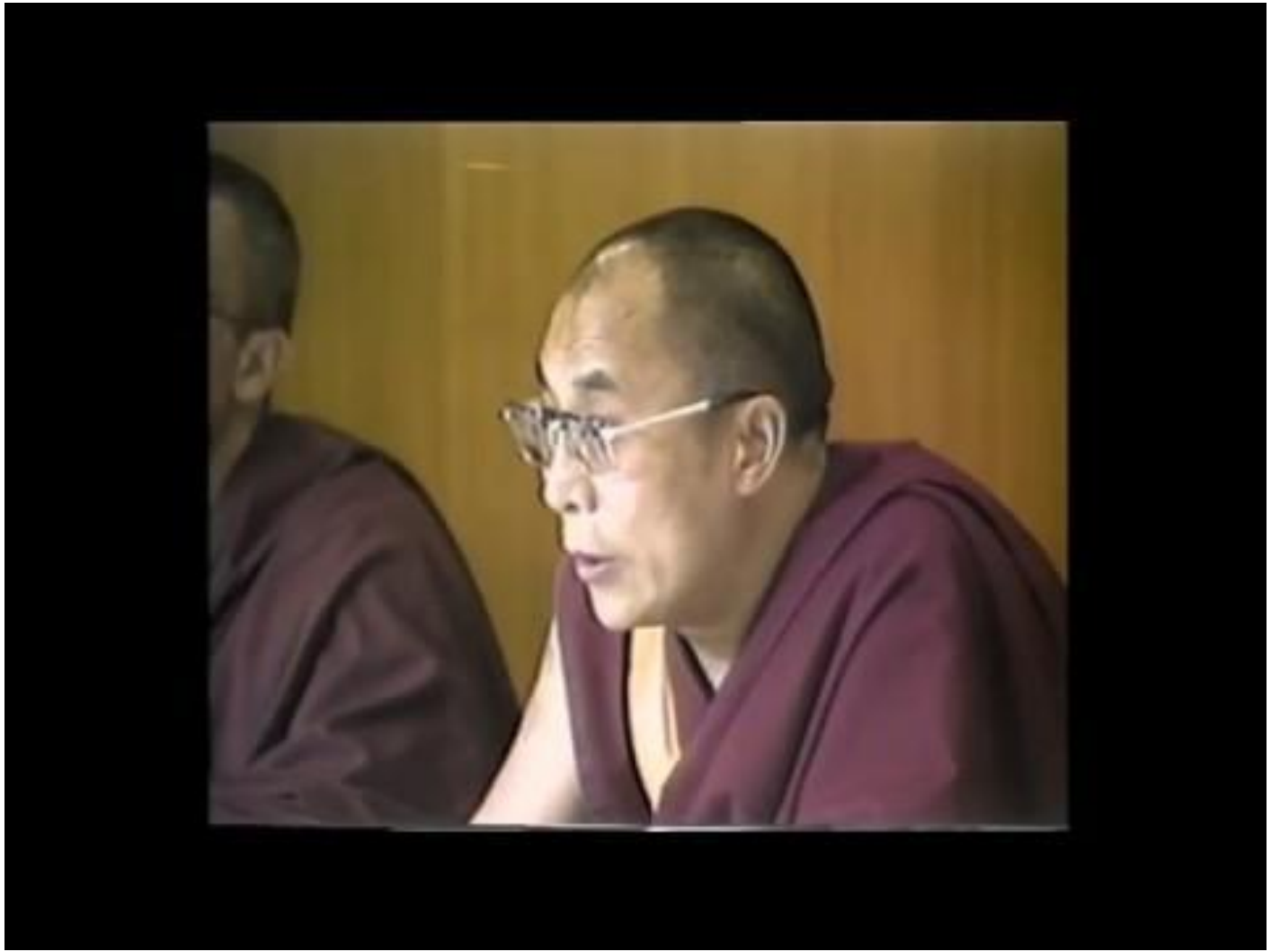
Human captioning ?

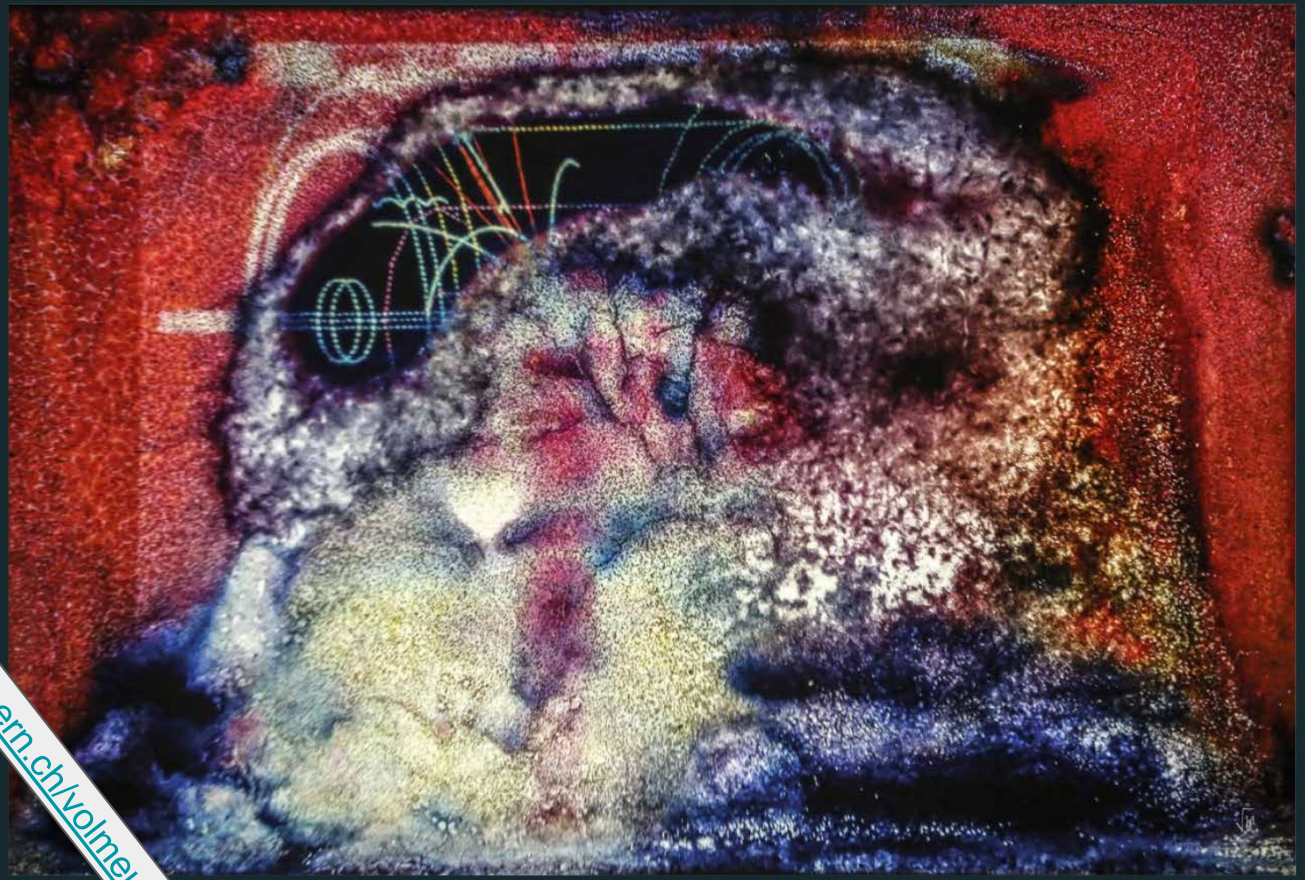
Quality Control

Surprise!

Surprise!

Surprise!





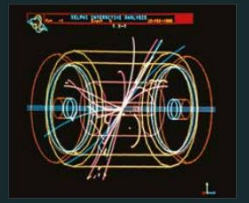
Breaking the mould

AT THE Large Hadron Collider near Geneva, Switzerland, physicists are used to looking for signs of particle decay in the detectors. But as they were digitising archival photos of particle collisions, Matteo Volpi and Jean-Yves LeMeur came across a different kind of decay: mould.

For 30 years, this slide was exposed to a mould that marched across the image, eating through the protein in the gelatin-based emulsion. The resulting chemical reactions left a chaotic swirl of colours and textures reminiscent of an abstract painting. To save the corroded image as it is now, Volpi and LeMeur shone a light through the slide and then photographed the projection.

Volpi is a photographer himself, and has tried to recreate the effect. "I've tried burning and freezing, and I use yeast and beer to create mould. It makes a nice effect, but I can't reproduce these colours and textures. I don't have 30 years to wait, like this mould did."

The slide was unearthed in a dusty desk drawer in a corridor of the experimental physics department at CERN. Like its better-preserved companion slide below, it showed a simulation of an electron-positron collision at DELPHI, one of four detectors at the LHC's predecessor, the Large Electron-Positron Collider. The blue horizontal lines represent the beams of particles that meet head on in the detector's cylindrical cavity, and the spray of arcs extending from the middle track the particles born in the smash-up. Chelsea Whyte

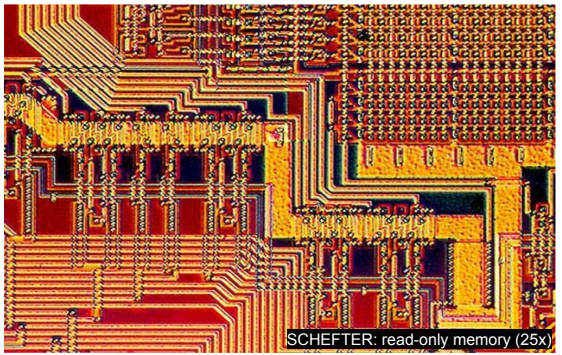


Photographer
Volmeur © 2017 CERN

http://cern.ch/volmeur

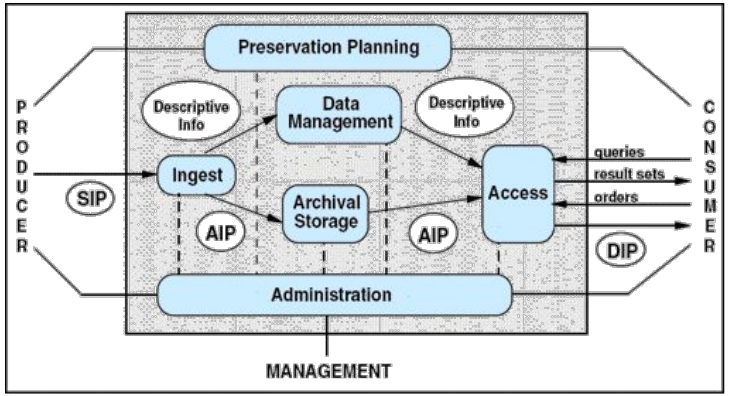
Entering the digital preservation paradigm

- All multimedia soon fully digitized (by mid 2019)
  - ~350’000 master & access files → up to 3’150’000 files (with all the formats, replicas and md5)
  - ~ 550 Tb : < 3% of CERN Data Cloud
  - retiring the good old tapes
- Facing the new risks
  - Physical obsolescence
    - bit rot; redundancy failure
  - Time obsolescence
    - readers, formats, OS, HW
  - Human obsolescence
    - lost in transitions; missing context; dissipation
    - economic failure; corruption by mistake or attack

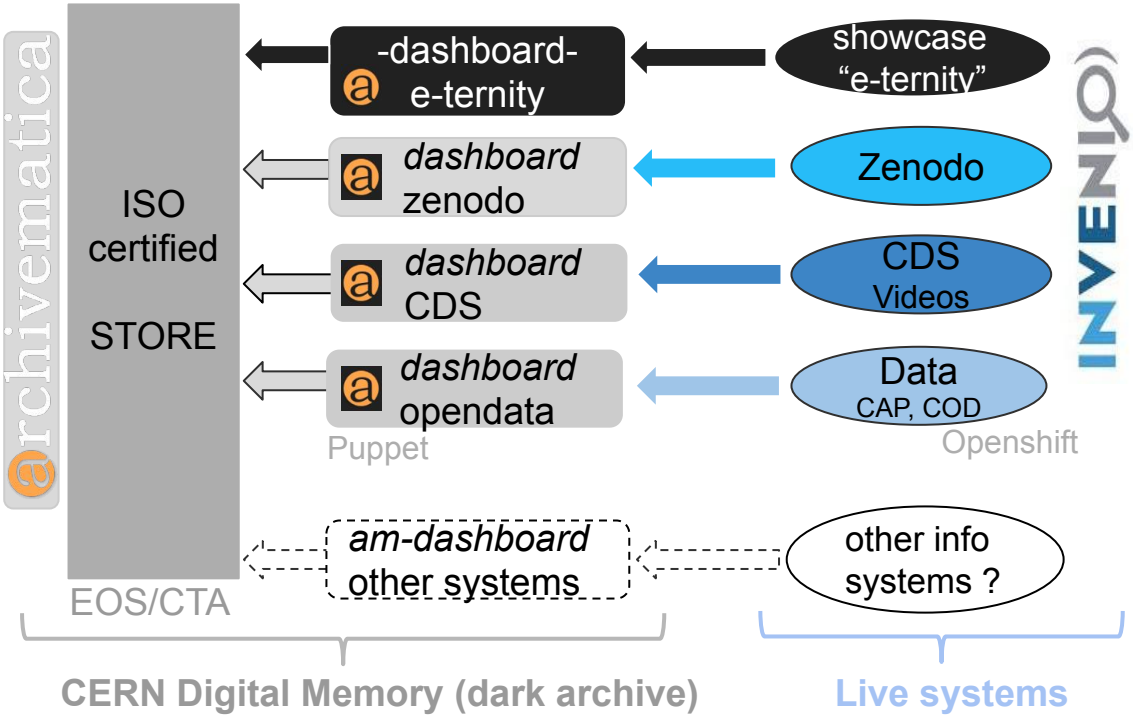


The Open Archival Information System (OAIS)

- Strict and powerful reference model
  - Trustworthy Digital repo “ISO 16363”
- Three levels
  - Organizational
  - Infrastructure
  - Digital Object
- AIP: Archival Information Packages
  - Self-sufficient
- Supported by existing SW
  - Preservica, Rosetta, Archivematica, etc
    - Creation of AIP (METS & PREMIS schemas)
    - Conversions to master formats
    - Workflow support



Integration of CERN Information Systems into an ISO 16363 OAIS Platform



ISO 16363

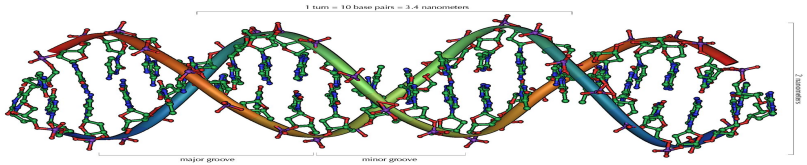
- Requirement of the European Strategy for Particle Physics (2013) → DPHEP
- On-going self assessment for the 2020 ESPP Update

Planned Developments

- Consolidating the showcase
- Extending the platform
  - Infrastructure (Db/ES-od)
  - Invenio-AM gateway
- Pluging Invenio digital repos
- Adding other systems
  - AFS phased out content

Concluding words

- Preserve the past
  - Digitize analog material
  - Archive it with digitally born content
  - On an ISO-certified platform on CERN cloud
  - Using Invenio & Archivematica OS SW
- Prevent “leak of memory” in the future
- Enlight the “dark archive”
  - Share Multimedia with grand public
  - Modern displays for old content
  - Wide interest in CERN historical content



Memory Collider:
Displaying Science